# ON THE EXPERIMENT OF BRUGMANS, HEYMANS, AND WEINBERG

Sybo A. Schouten
University of Utrecht

Edward F. Kelly
Duke University

Beginning in May 1920 an ESP experiment was carried out by Brugmans, Heymans and Weinberg of the University of Groningen, the Netherlands, with a gifted subject, a student named van Dam. This experiment, of which some early results were reported by Brugmans at the first and second international conference on parapsychology in Copenhagen and Warsaw, raised much interest both because of the exceptional performance of the subject and because of the many interesting features of the study. Of special historical interest is the fact that psychophysiological methods were applied to study the relationship between the 'state' of the subject and his ESP performance, which makes this experiment probably the first in parapsychology in which psychophysiological variables were studied.

The reports of the experiment that have previously been available in French and English (for example, Brugmans (1922, 1924), Carington (1946), Hansel(1966), Murphy(1927,1961), Pope(1952), Soal and Bateman (1954)) are terse and fragmentary. The results which Brugmans (1922) reported were based only on the data of the first seven sessions comprising some 187 trials held between May 1920 and September 1920. In fact, the experiment lasted till June 1922 and ultimately consisted of 24 sessions and 589 trials. Apparently the experimenters never felt it worthwhile to report the results of the additional sessions, thinking that by the end of the experiment van Dam had lost his ability. As we will show, even at the end he was doing quite well by contemporary standards.

The experiment was in several respects unusually sophisticated for its time, and for a number of years afterward was widely regarded as one of the more substantial bodies of evidence supporting the existence of psi phenomena. In recent decades, however, its reputation seems to have slipped considerably. In part this no doubt reflects the normal erosion through neglect that comes with passage of time. The main factor, however, is certainly a series of seemingly damaging critical attacks on the experimental procedures.

The most serious such attacks involve two categories of possible problems: First, it has been suggested that the target orders were so structured as to render statistical analysis of the direct hit results invalid (for example, Rhine, 1977, p. 28; to our knowledge, no supporting evidence for this claim has ever been published). Second, the suggestion has been made by Soal and Bateman (1954)--and eagerly accepted by Hansel (1966)--that van Dam's entire performance might be the result of sensory leakage.

We have been fortunate enough to gain access to the original targets and response series of the entire experiment. Also available, in Dutch, is a series of reports in the "Meededeelingen der Studievereeniging voor Psychical Research" (Proceedings of the Dutch Society for Psychical Research). Taking advantage of this opportunity, our aims in this report are several:

1) to provide a relatively complete English-language account of the experiment and its main results as regards targets, responses and scoring;

2) to redo the main analyses using modern methods, and to extend these analyses to the previously unreported data;

3) to address specifically the critical questions raised above concerning target orders and sensory leakage; and

4) to report a variety of novel analyses of the complete data, including both new analyses directed to old questions and analyses for previously unreported effects.

In a following publication we hope to be able to present more data on the results of some variables studied in these experiments, especially those concerning the psychophysiological measurement of the 'passieve staat' of the subject and those concerning the employment of drugs like alcohol and bromium.

## Description of the experiment

In the summer of 1919 the experimenters, Prof. G. Heymans, Dr. H.J.F.W. Brugmans, and A.A. Weinberg, came in touch with a student in mathematics and physics, A.S. van Dam, 23 years old, who felt that he was telepathically gifted and who offered himself as a subject for investigations. Heymans was at that time director of the first Dutch Laboratory for Experimental Psychology, Brugmans one of his staffmembers, later to become Heyman's successor, and Weinberg the conservator of the laboratory.

The first pilot studies with van Dam were carried out by Weinberg and van Loon. Van Loon was probably another staffmember who shortly afterwards left for the Indies and therefore could not participate in the main experiment. The results of these pilot studies which involved in addition to van Dam two other subjects, were reported

in the Journal of the Society for Psychical Research (January & February 1921). In these experiments various sensory modalities were employed as target objects. It was tried to transfer telepathically colours, tastes, feelings and moods, because the experimenters felt that "very probably - not all the contents of consciousness could be transferred with equal ease, considering the fact that the various processes of consciousness make a very different impression on the human mind". Furthermore they were interested whether one kind of mental process especially would be transmitted more easily than others, i.e. impressions of an emotional character.

From the results of these experiments the authors concluded that "our experiments undeniably demonstrate that extra-sensorial transmission of the contents of consciousness is possible" (J.SPR February 1921, p.44). Another conclusion was, that emotional processes of consciousness are more easily transmitted than others, and that the impressions received by the best developed organs of sense were very likely transmitted most easily. The latter conclusion has undoubtedly played an important role in the decision to employ the transfer of motoric impressions in the main experiment with van Dam.

Unfortunately, because of the design these pilot studies do not lend themselves for statistical analyses. Another objection is that all sessions were held with agent, subject and observer being in the same room. Apparently Heymans and Brugmans were therefore not so convinced of the validity of the results as were Weinberg and van Loon. For the main experiment they took extensive precautions to avoid sensory leakages, like blindfolding the subject, putting agent and subject in different rooms,etc. Secondly Heymans et.al. stated in his report to the Dutch SPR (1921) that pilot studies with van Dam were carried out, involving the transfer of taste, colour and forms, but that the results although being positive could not be considered as striking. It should be noted, however, that Heymans and Brugmans had clearly a different opinion about what constituted a 'striking result' than we have nowadays. They seem to have applied a rather subjective criterion which required a much higher scoring rate than would presently be required by criteria based on statistical significance.

van Dam showed a strong initial preference for what was termed in those days "Rubini" experiments. In such an experiment (directly modeled after stage performances by an individual of that name) an object is hidden and the subject, moving around freely, tries to locate it while someone knowing the place of the object remains in the vicinity acting as agent. The experimenters recognized

clearly that all such procedures, including those involving no direct contact between agent and subject, were defective because of various possibilities of sensory leakage. They therefore set out to design a procedure which would be free of these defects, while retaining maximal resemblance to the "Rubini-type" situation, which van Dam enjoyed. (Rubini-type trials were still employed during later experimental sessions, but only as a kind of warm-up for the formal testing to be described; the results of these trials were apparently extremely successful, but they were strictly segregated from the results of the formal trials). An interesting point arose when the experimenters discovered that it was very difficult for van Dam to visualize. Instead he relied heavily on motoric imagery. Thus for instance forms were represented internally as movements. Hence the experimenters decided to design their experiment in such a way that van Dam would be able to apply this ability in a rather direct and congenial way. The central idea of the experiment was that agent-experimenters would try to transfer "telepathically" motoric impressions to van Dam. It is important to note that the experiment was thus not designed as a guessing task. Rather, the object was to transfer impressions of movements.

In order to be able to evaluate the results of the attempted transfer of these impressions, a sort of checkerboard was constructed with 48 squares. This board measured 40x30 cm., each square being 5x5 cm. The squares were indicated by giving the eight columns the letters A to H (from left to right) and the six rows the numbers 1 to 6 (from bottom to top). The letters and numbers were printed on the board. Each square of the bottom row contained the letter of the corresponding column, and on each square of the other five rows the number of the row was printed. According to some photographs printed in the "Meededeelingen" the experimenters had a clear view of the letters and numbers from their position in the other room (see also Murphy, 1961). We do not know whether it was possible for van Dam to distinguish the boundaries between the squares, when moving his hand over the board.

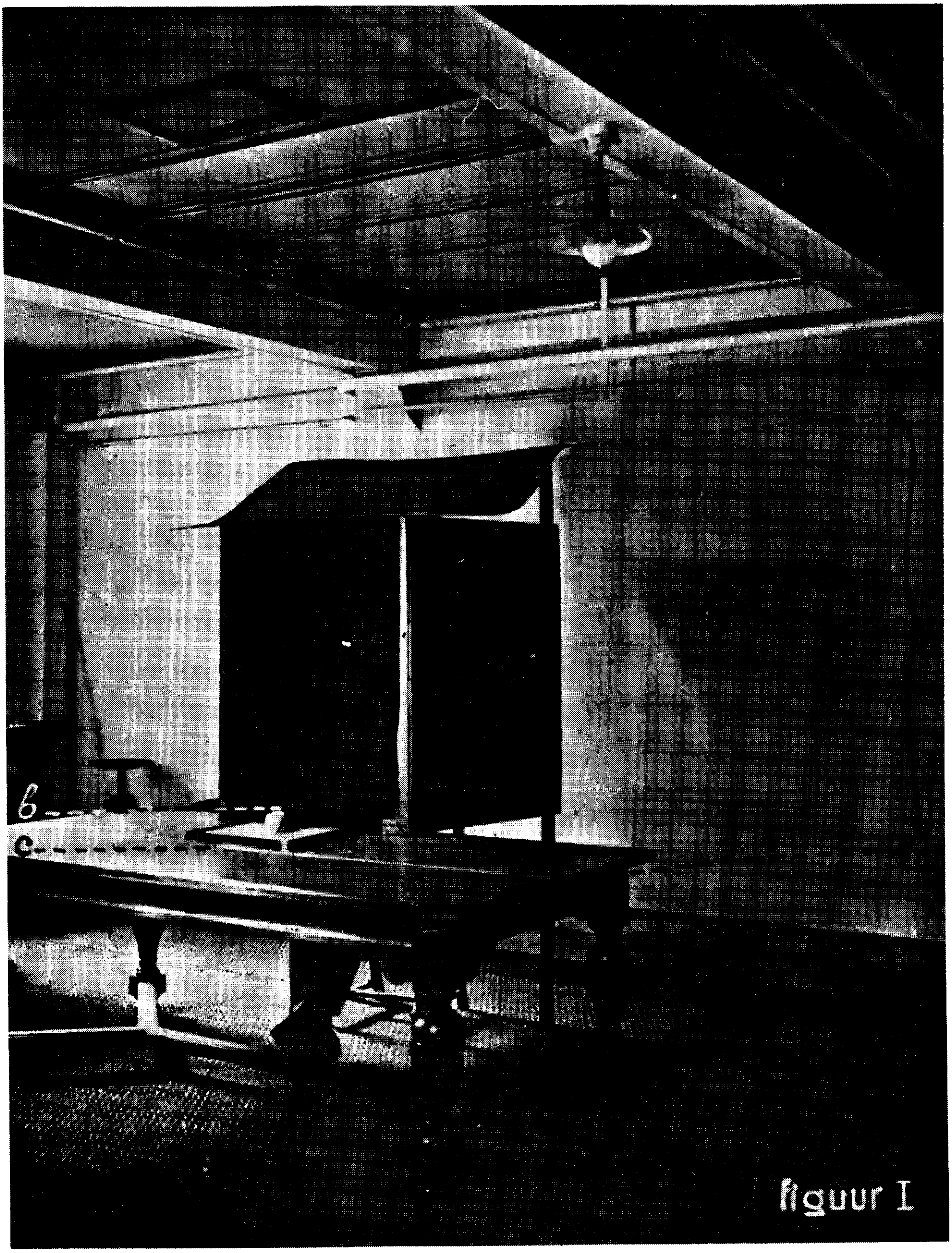
This checkerboard was chosen because it would enable the experimenters to fix a certain point as the target area for the motoric movement, and because they would be able this way to fix the probability of a direct hit. The task of the agent was to 'steer' the hand of the subject to the target area (in a Rubini experiment the location of the hidden object), by transferring the impression of the movement of the hand from the present position to the target area. If the subject felt that he had reached the target area, he indicated this by tapping his finger twice on that

area. It was stressed, however, that the subject should only give the agreed-upon signal when he felt that he had reached the target area. In other words, the subject was in fact required not to guess, but to wait till he received the impression that his finger had reached the target square.

So actually the experiment was more complicated in its psi aspects than it looks superficially. In fact each trial consisted of a varying number of "subtrials" because every time the subject shifted the position of his hand, the agent started a new subtrial by trying to transfer the movement from the new position of the hand to the target area. Secondly, after successfully transferring these movements the agent also had to transfer the impression that the target area had finally been reached and that the subject should stop moving his hand. From the report in the "Meededeelingen" it can be concluded that van Dam clearly experienced the latter as a separate task. He had to receive a distinct impression that his task for that trial was finished.

Hence a "hit" can be thought of as terminating an unknown number of subtrials, in which momentary motoric impressions were to some extent successfully transferred, and in which at a given moment the impression was also successfully transferred that the task was completed and that the subject should give the signal. A trial was rated as a miss, if the subject ended the trial by giving the signal when pointing to a square different from the chosen target square. But even in these trials an unknown amount of successful transfer of motoric movements might have occurred. In fact, the authors state in their report that occasionally it happened that in a given trial the hand was repeatedly 'steered' to the target square, but that the agent failed each time in transferring the impression that the target area had been reached. We discuss this so extensively here because we feel that it is important to realize that while the statistical analysis presented is ultimately based only on whether the target area was correctly identified or not, this aspect-the feeling of van Dam that he had reached the target square-constituted only part of the experiment.

The main psi aspect, the transfer of the motoric impressions, can not be analysed because we lack detailed information both about van Dam's hand movements and about the specific impressions which agents tried to convey. Because of the fact that the main part of the experiment lies in the 'steering' of the hand, however, it might be expected that the behavior of the subject as far as the 'selection' of the response squares is concerned might have been rather different had he been involved in a pure guessing task. Some statistical evidence consistent with this expectation


b
c
figuur I

is presented below.

The experimenters took a number of special precautions to exclude possibilities of sensory leakage. Two rooms were used for the experiment. In one of the rooms in a sort of half-basement of the Psychological Institute a cabinet was erected in which van Dam was seated. Three sides of this cabinet were closed, the sides left and right of the subject and the top. The side in front of the subject was covered with a heavy curtain in such a way that he could pass his hand under the curtain and move it on the surface of a table in front of the cabinet on which the checkerboard was put. This cabinet was applied to ensure that he could not see the ceiling of the room or the checkerboard on which his hand moved. As a further precaution, and also to help him more easily reach a relaxed state, van Dam remained blindfolded during the session.

In the ceiling of this room, which judging from the photographs was rather high, an opening was made of 32x52 cm. The separating floor between the two rooms, the upper one located on the ground floor of the building, was, according to Carington (1938) who inspected the rooms personally, of 'a most uncompromising solidity'. The floor of the upper room consisted of planks of about 2,5 cm. thickness, laid on about 25 cm. thick joists which were spaced at about 35 cm. The ceiling of the basement room consisted of lath and plaster fixed to the joists. The opening mentioned above was sealed off at both sides with glass windows. The distance between the glass windows must have been approximately equal to the thickness of the joists, that is about 25 cm. Although according to the photographs in the "Meededeelingen" the floor of the upper room was normally not covered with rugs, a thick mat was used on which the agent stand during the trials.

The sound isolation between the two rooms was of such quality that according to the report in the "Meededeelingen" the experimenters could not make themselves understandable from one room to the other even when shouting at the top of their voices. Hence when the experimenters stayed in the upper room, the onset of a trial was indicated to van Dam by striking the floor with a hammer. During the sessions the upper room where the experimenters stayed was kept dark by closing the curtains and shutting off all lights.

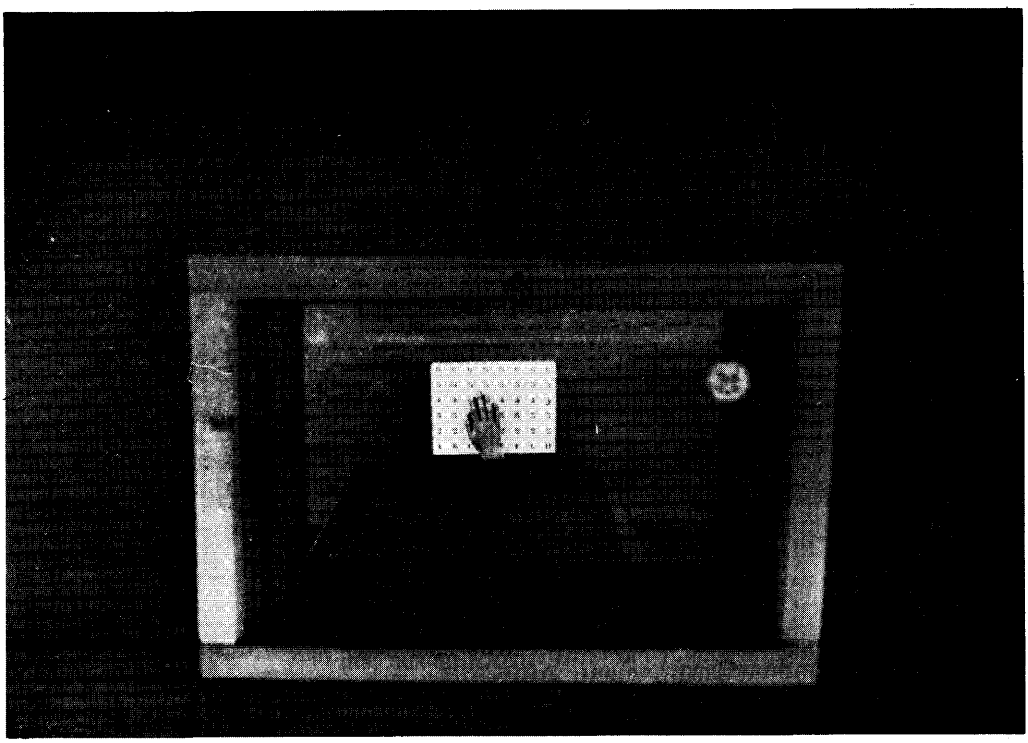
In each trial one of the experimenters acted as the agent, while the others functioned as observers. They changed roles within a session in a systematic way. According to the report, the agent selected the target square and the observers were not explicitly informed about the target until after completion of the trial.

The agent selected the target square by shuffling separately two stacks of cards, one containing the letters A to H, the others the numbers 1 to 6. Each stack contained one copy of each card.* After shuffling, the agent mentally noted the target and placed the two stacks upside down on the table. The bottom cards facing the surface of the table represented the letter and number of the target square. After completion of the trial one of the observers picked up the stacks and recorded the letter and number. It is not stated clearly when this recording took place, but it can be assumed in view of the whole procedure that it happened after the response of the subject was recorded by the observers. On the other hand, it seems possible that the observers might get some idea of the location of the target area during the trial, based on the behavior of the agent as he attempted to influence the movement of the hand of the subject. The reports do not discuss this possibility, but in the "Meededeelingen" it is stated explicitly that the observers only learned the identity of the target after the trial, when it was recorded.

In the event the subject tapped in between two squares, both possible response squares were recorded whether the ambiguity was in letter, in number, or both. The experimenters invariably rated trials with an ambiguous response as a miss, including those trials in which one of the recorded response squares was identical with the target square. On all but 17 trials the response time was also recorded. Although this is not specifically stated in the report, we can assume that at least in the distance condition, where all experimenters stayed in the upper room, the subject was not informed after each trial about the details of that trial. It is not likely that one of the experimenters went down the staircase and back again only to tell van Dam how well he succeeded or what the target was. It is more likely that van Dam was informed about the results after the session was completed. At least as far as the targets are concerned this assumption is backed up by the statistical analyses, presented below.

It is also not clearly stated which hand was used by the subject when making the movements. In the "Meededeelingen" three pictures are shown of the cabinet with a person sitting in it. In all three pictures a right arm appears from under the curtain. Assuming that these pictures were taken sometime during the sessions, the subject can be supposed to have used his right hand.

* In some previous reports (for example Hansel (1966)), the procedure is incorrectly described as one in which slips of paper bearing the letters and numbers were drawn from bags.



In total 589 trials were run, in 24 sessions of unequal length. (For two trials near the end of the experiment no response was recorded; these trials were excluded from the analyses reported below). The experimenters allocated trials in varyingly systematic ways to conditions whose effects on the performance they wished to study. These include in particular experimenters, distance, and certain physiological factors. The three principal experimenters were Brugmans, Heymans, and Weinberg, although a fourth, currently unidentified experimenter contributed eight trials near the end of the experiment. Distance was studied at two levels, a distant condition (233 trials) in which all experimenters remained in the upper room watching van Dam through the porthole in the ceiling, and a near condition in which the agent took a position about 1 meter away from the checkerboard in front of the subject. It is not clearly stated where the observers stayed in these trials, but we assume that they stayed in the upper room. The experimenters were aware that in the near condition the possibility of some sensory cueing could not be excluded. Although the agent tried to control himself as much as possible, they state that: "in the near condition tactile and visual impressions were completely excluded, but auditory cues, e.g. because of movements and respiration, could of course not be considered to be entirely excluded" (Meededeelingen, p.6). Physiological aspects of van Dam's performance were studied in two main ways; by including some sessions in which he was given either alcohol or bromide in advance, and by attempting in some sessions to quantify his trial-by-trial physiological state using measures of galvanic skin response, pulse, and respiration. As stated before we hope to discuss these aspects in a subsequent publication on this experiment.

## RESULTS

### Analysis of the Target Sequence

As mentioned above, questions have been raised about the adequacy of the target sequences obtained. In this and succeeding sections we give detailed information sufficient to answer these questions. For the analysis of the target and response series the whole sequence of trials is taken as one unit. We assume that possible non-random effects in the target series are most likely due to inadequate shuffling of the decks of cards. Since the decks must have been specially constructed for this experiment, they probably were left undisturbed between the sessions.

## Zero-order target effects

The first question is whether the possible targets were selected with equal frequency apart from expected statistical fluctuations. The overall frequencies of the target letters, numbers, and individual squares are presented in Table 1.

TABLE 1

Frequency distributions for target letters, numbers, and individual squares

| | | Number | | | | | | Letter totals |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | |
| Letter | A | 15 | 20 | 21 | 13 | 25 | 9 | 103 |
| | B | 18 | 16 | 20 | 19 | 13 | 16 | 102 |
| | C | 12 | 13 | 16 | 12 | 17 | 4 | 74 |
| | D | 6 | 11 | 9 | 7 | 5 | 5 | 43 |
| | E | 9 | 13 | 17 | 6 | 18 | 10 | 73 |
| | F | 11 | 13 | 13 | 10 | 13 | 11 | 71 |
| | G | 12 | 22 | 10 | 12 | 31 | 11 | 98 |
| | H | 2 | 6 | 4 | 3 | 4 | 4 | 23 |
| Number totals | | 85 | 114 | 110 | 82 | 126 | 70 | 587 |

It is evident from these data that the shuffling procedure was not adequate to ensure equal sampling probabilities for letters, numbers, or squares.

For letters, $\chi^2 = 78.6$, df = 7, $p \ll .001$. For numbers $\chi^2 = 24.5$, df = 5, $p < .001$. There is systematic over-representation of the letters A, B and G and the numbers 2, 3, and 5; and systematic avoidance of the letters D and H and the numbers 1, 4, and 6. Not surprisingly, the frequency distribution for individual squares is correspondingly non-uniform, with $\chi^2 = 145$, df = 47, $p \ll .001$.

An important question is whether the processes of selection for numbers and letters were statistically as well as procedurally independent. This is easily tested by computing the chisquare for cell frequencies, given the marginal frequencies. The resulting value is 32.1 with 35 degrees of freedom, which is entirely non-significant. This result strongly supports the experimenters' statement that the target squares were chosen by independent selection of their letter and number attributes.

The pattern of non-uniformity in zero-order frequencies for letters and numbers proved to be similar for near and distant conditions (Table 2) as well as for ambiguous and non-ambiguous trials (letters, $\chi^2 = 8.8$, df = 7; numbers, $\chi^2 = 4.4$, df = 5).

TABLE 2

Target letters and numbers in near and distant conditions

| Letters: | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| near | 62 | 61 | 41 | 29 | 40 | 46 | 62 | 13 |
| distant | 41 | 41 | 33 | 14 | 33 | 25 | 36 | 10 |

$\chi^2 = 3.7$ df = 7

| Numbers: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| near | 52 | 70 | 70 | 51 | 68 | 43 |
| distant | 33 | 44 | 40 | 31 | 58 | 27 |

$\chi^2 = 2.9$ df = 5

It was also fairly stable across the duration of the experiment: Dividing the experiment into halves and comparing target frequencies for letters and numbers in the two halves, we find no difference for numbers and only a marginally significant difference for letters (see Table 3).

TABLE 3

Frequency distribution of target letters
First vs. second half of experiment

| letter | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| first half | 50 | 38 | 36 | 21 | 34 | 46 | 52 | 16 |
| second half | 53 | 64 | 38 | 22 | 39 | 25 | 46 | 7 |

$\chi^2 = 17.2$ , df = 7 , p~ .02

The main changes contributing to this result are the increased frequency of B and the decreased frequency of F in the later part of the experiment. Note too that the overall pattern is if anything more non-uniform in the second half.

So far it appears that the non-uniformities of the target frequencies may be uniquely associated with physical properties of the decks of target cards. However, there is also apparently a contribution from the experimenters, possibly related to their shuffling techniques. The frequencies of target numbers do not differ across the three main experimenters, although there is a trend in this direction ($\chi^2 = 17.8$, df = 10, $.05<p<.10$); however, for letters there is a strongly significant effect (see Table 4).

TABLE 4

Experimenters vs. target frequencies for letters*

| letter | A | B | C | D | E | F | G | H | Totals |
|---|---|---|---|---|---|---|---|---|---|
| Brugmans | 26 | 71 | 34 | 8 | 25 | 27 | 29 | 8 | 228 |
| Heymans | 36 | 19 | 26 | 18 | 21 | 19 | 30 | 6 | 175 |
| Weinberg | 38 | 11 | 13 | 16 | 26 | 25 | 38 | 9 | 176 |
| Totals | 100 | 101 | 73 | 42 | 72 | 71 | 97 | 23 | 579 |

* unidentified fourth experimenter omitted

$\chi^2 = 69.5$ , df= 14, $p<<.001$

The most conspicuous contributors to this result occur in the column corresponding to 'B' targets, which are strongly over-selected by Brugmans, and less strongly under-selected by Heymans and Weinberg. Nonetheless there is still some visually obvious evidence of uniformity, particularly in the low frequency of 'H' targets. Furthermore, a statistical measure of agreement in the ranked frequency distributions across experimenters is still almost significant (Kendall's W = .61 corrected for ties, $.05<p<.10$; Siegel, 1956). So although there is evidence of disparity, particularly in the frequency with which the different experimenters selected letter attributes of target squares, those differences are not too large, and leave considerable evidence of some more basic selection biases originating presumably in the physical properties of the target cards. Fortunately, as we will show in the sequel, the scoring rates are virtually identical across experimenters.

First-order target effects

First-order sequential dependencies were analyzed using a test which is asymptotically independent of the observed zero-order biases (David and Akers, 1974). For letters the analysis yields $\chi^2 = 67.1$ with 49 df, $p \sim .05$, 1-tailed, suggesting weak dependencies among successive letter targets. No such effect was found in the target number sequence ($\chi^2 = 19.5$, df = 25).

A sequential analysis on the complete targets is not possible because of the too low frequencies of most of the cells. Instead a "distance" analysis was carried out between the successive targets. Distance is defined here as the square root of the sum of the squared differences in number of letters and numbers between two squares. For instance, between A1 and C5 the distance is SQRT($2^2+4^2$). In total there are 31 possible different distances between squares. Such an analysis can detect different kinds of sequential dependencies which are not related to individual targets-- for instance, a tendency to select the next target relatively far away from the preceding one. The expected distribution is obtained by enumeration of all the possible distances between all the targets. (The actual target set was applied, because it had already been shown that the targets were not selected with uniform probability). Comparing the expected distribution of distances with the distribution based on the actual target sequence (see figure 1) no difference was observed ($\chi^2 = 23.5$, df = 30). Applying this same analysis to the observed distribution of successive distances for target letters and numbers separately also leads to non-significant results (letters: $\chi^2 = 5.0$, df = 7; numbers: $\chi^2 = 5.5$, df = 5). This implies that the marginally

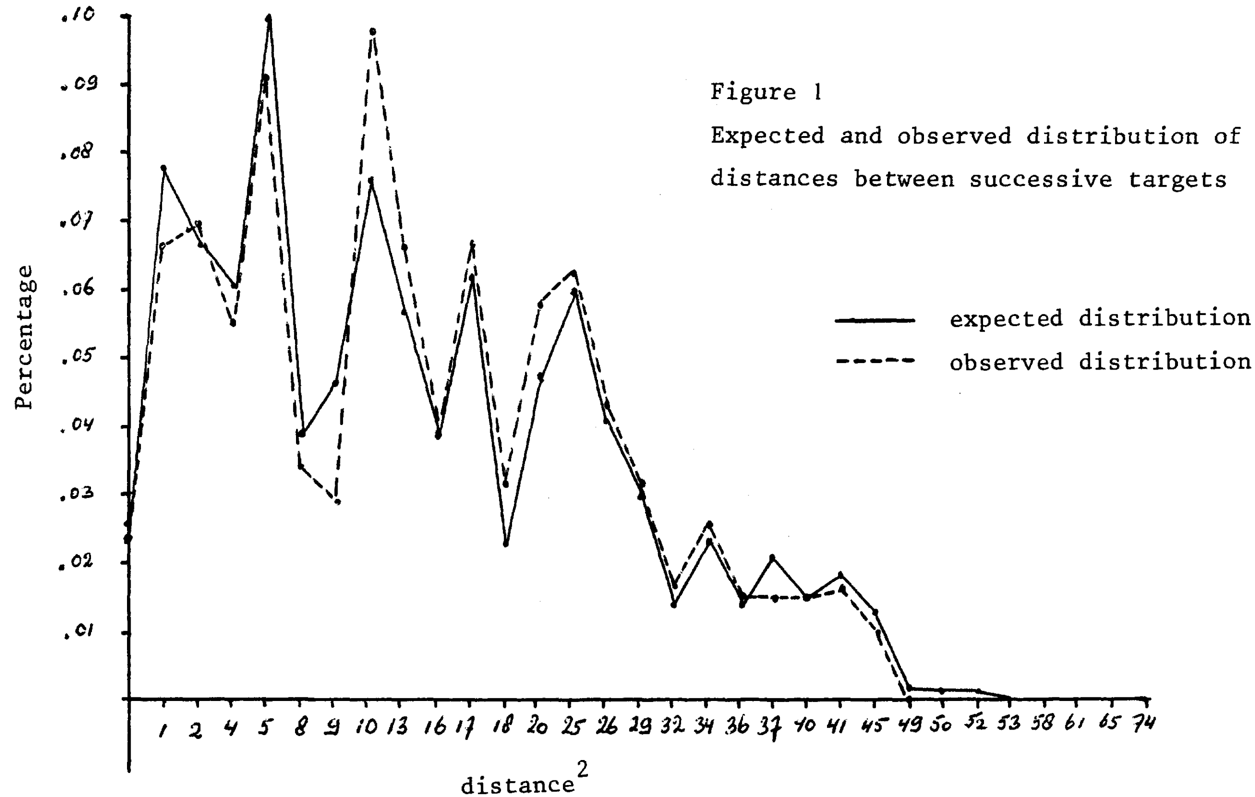



Figure 1

Expected and observed distribution of distances between successive targets

significant sequential dependency in the target letter series cannot be based on a general tendency affecting all letters. On the other hand, a study of the differences between the observed and expected frequencies of the sequential target letter combinations did not reveal any systematic pattern, nor could we find a specific set of combinations which can be held responsible for this marginal significant effect.

Some additional analyses have been carried out to check obvious candidates for specific kinds of non-random sequential effects. For example, dividing the checkerboard into left and right halves, the sequential target combinations were transformed into left-right combinations to see whether there exists a systematic tendency to change the targets from one half of the board to the other. No such effect was found (see Table 5).

TABLE 5

Sequential targets: left-right

| | trial N+1: | left | right | |
|---|---|---|---|---|
| trial N | left | 172 | 149 | $\chi^2 = 0.3$ df = 1 |
| | right | 149 | 116 | |

The same analysis was applied dividing the board into upper and lower halves (see Table 6).

TABLE 6

Sequential targets: up-down

| | trial N+1: | up | down | |
|---|---|---|---|---|
| trial N | up | 129 | 147 | $\chi^2 = .03$ df = 1 |
| | down | 148 | 162 | |

Note: all 1 df chi-squares corrected for continuity

An equally non-significant result was observed when dividing the board into a central part and a side part (see Table 7).

TABLE 7

Sequential targets: central-side

| | centre | side | |
|---|---|---|---|
| centre | 184 | 139 | $\chi^2 = 0.8$ df = 1 |
| side | 139 | 124 | |

Finally, in an effort to avoid overlooking any previously unsuspected but strong forms of sequential dependency in the targets, we traced out manually some of the successive transitions between targets, separately for numbers, letters, and squares. This was done for approximately the first 100 trials in each case, carrying us well past the period of van Dam's most dramatic scoring (sessions 2 and 3). No new candidates were discovered in this way.

From all these analyses we conclude that there is little evidence of sequential dependencies among successive targets, except for the very marginal and diffuse effect involving individual target letters. Moreover, no systematic pattern could be found which might explain even this very weak dependency; hence it is difficult to see a priori how the subject could have taken advantage of this effect even had he known about it. Since the targets were generated successively, and since the distance analysis did not yield any significant deviations for numbers, letters, or squares, higher-order effects appear even more unlikely.

Analysis of the Response Sequence

In 131 cases the observers could not discriminate between two adjacent squares when the subject tapped his finger, and wrote down both squares as response. Murphy (1961) appropriately raises the question whether the manner in which response ambiguities are resolved has any impact on the main statistical results. In order to bracket the possibilities we decided to create and analyze in parallel two versions of the data, in each of which the ambiguous responses are uniformly resolved in one of the two possible extreme ways. Thus one dataset was constructed by always choosing for ambiguous responses the more distant response from the target square (the MDR set); while for the other set (the LDR set) the less distant response was uniformly chosen. (For non-ambiguous responses, of course, the datasets are identical). Unless

otherwise stated, all analyses reported below were carried out on the MDR set. As it happens, none of these statistical results differ in more than minor respects between the two sets. Most analyses have been carried out independently by the two authors. Afterwards it turned out that the MDR and LDR set of responses used by us differed slightly due to the fact, that in a few ambiguous trials both responses had an equal distance to the target square, in which case each of us randomly assigned one of these responses to each set. This also created some minor discrepancies between different tables presented below.

Zero-order response effects

The overall frequencies of the response letters, numbers, and squares are presented in Table 8.

TABLE 8

Frequency distributions for response letters, numbers, and individual squares

| | | Number | | | | | | Letter |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | totals |
| Letter | A | 5 | 10 | 12 | 11 | 15 | 3 | 56 |
| | B | 9 | 10 | 12 | 20 | 9 | 5 | 65 |
| | C | 9 | 17 | 27 | 14 | 10 | 5 | 82 |
| | D | 18 | 12 | 21 | 24 | 15 | 3 | 93 |
| | E | 7 | 7 | 21 | 43 | 14 | 8 | 100 |
| | F | 8 | 11 | 16 | 24 | 15 | 11 | 85 |
| | G | 7 | 13 | 20 | 16 | 16 | 4 | 76 |
| | H | 3 | 3 | 3 | 7 | 11 | 3 | 30 |
| Number totals | | 66 | 83 | 132 | 159 | 105 | 42 | 587 |

As was the case for the targets, there are strong zero-order response effects at all levels: For letters, $\chi^2 = 48.6$ with 7 df, p<<.001. For numbers, $\chi^2 = 95.2$ with 5 df, p<<.001. For squares, $\chi^2 = 227$ with 47 df, p<<.001. Furthermore--and in contrast to what we saw for targets--the response attributes are not chosen independently; van Dam's choices of letters and numbers are significantly associated ( $\chi^2 = 57.9$, df = 35, p<.01). From inspection of Table 8 it is apparent that he had a strong general tendency to terminate the trial on more centrally located squares.

These response patterns are rather stable throughout the experiment. They are consistent across such factors as near vs. distant conditions (letters, $\chi^2 = 15.3$, df = 7, p ~ .05; numbers, $\chi^2 = 6.8$, df = 5), clear vs. ambiguous responses (letters, $\chi^2 = 10.8$, df = 7; numbers, $\chi^2 = 9.3$, df = 5), and experimenters (letters, $\chi^2 = 14.6$, df = 14; numbers, $\chi^2 = 8.2$, df = 10). (For near vs. distant, the letters are chosen with more nearly uniform frequency in the distant condition).

When the frequencies of response letters and numbers are compared for first and second half of the experiment, the number frequencies are again consistent, but there is again significant difference in the frequency distribution of the letters (see Table 9). Since distant trials were more frequent in the first half of the experiment, this effect is correlated with the one just mentioned, and of the same form--i.e., letters are chosen with more uniform frequency in the first half.

TABLE 9

Frequency distribution of response letters
First to second half of experiment

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| first | 30 | 31 | 37 | 38 | 42 | 50 | 46 | 18 |
| second | 26 | 33 | 45 | 54 | 60 | 33 | 31 | 12 |

$\chi^2 = 14.6$ df = 7 p = .05

We shall say more about this table subsequently in discussing the question whether van Dam's response patterns show any evidence of reflecting the biases of the target sequence.

First-order response effects

No first order dependencies between the successive responses were observed for the response letter combinations ( $\chi^2 = 42.7$, df = 49), but a marginally significant first-order sequential dependency was observed for the numbers ( $\chi^2 = 41.8$, df = 25, $p<.02$). Note that this is opposite from what we found when analyzing the target series. The sequential dependency for the response numbers is not due to the familiar response bias to avoid repetitions. In fact, both in the number and letter response series the number of observed repetitions slightly exceeds the expected frequencies.

The distance analysis of the response series showed only non-significant differences between the observed distribution of distances of the successive response squares and the expected distribution ( $\chi^2 = 22.1$, df = 30). The same applies to the distribution of the distances between the successive response letters ( $\chi^2 = 3.8$, df = 7) and to the distribution of the distances of the successive response numbers ( $\chi^2 = 9.1$, df = 5).

Analyzing for specific non-random patterns in the response series, no effect was observed when comparing successive responses as regards left and right half of the board ( $\chi^2 = 1.31$, df = 1) or upper or lower half of the board ( $\chi^2 = 1.96$, df = 1). A slight effect was found when comparing successive responses marked as central or side squares (see Table 10), its form being a weak tendency to choose the next response in the same area (central or side) of the board as the preceding one.

TABLE 10

Sequential responses: central-side

| | trial N+1: | centre | side | |
|---|---|---|---|---|
| trial N | centre | 325 | 104 | $\chi^2 = 4.2$ df = 1 |
| | side | 105 | 52 | $p \sim .05$ |

We emphasize that these sequential dependency effects in the response series are unusually slight compared to the sequential dependencies which are typically found when subjects participate in a guessing task.

This confirms what we stated in the description of the experiment, that the experiment was not designed as a guessing task and was probably not experienced as such by the subject. It should also be noted that these effects are lacking in the target series.

Relationships between target and response biases

On the assumption that complete trial-by-trial feedback was provided, van Dam obviously had opportunity to acquire information about properties of the target sequence, whether consciously or unconsciously. Even if only hit/miss feedback or no feedback at all was provided, however, it is still important to inquire whether biases in the response data are in significant respects congruent with those of the target data.

We have addressed this question in a number of ways. First, we generated scatterplots and spearman rank-order correlations comparing zero-order frequencies in target and response data for letters, numbers and squares, and comparing first-order frequencies for letters and for numbers.

As indicated above, there is strong evidence of non-uniformity in overall selection frequencies for numbers, letters, and squares for both targets and responses (Tables 2 and 8). However, despite the conspicuously low frequency in both series of the letter H and the number 6, the overall relationship between target and response biases is for both attributes non-significant (for the letters it is, in fact, an inverse relationship; $r_s = -.275$, NS). Although for squares there is a marginally significant positive correlation ($r_s = .308$, $t = 2.19$, $df = 46$, $p<.05$) it is clear from the plots that this weak relation is dependent on the low frequency in both series of the compounds of H and 6.

The weak evidence of sequential dependency in the target and response series occurred only for letters (but not number) and numbers (but not letters) respectively. Moreover, the relationships between overall first-order frequencies of targets and responses for both letters and numbers are non-significant, although there is a trend for the numbers ($r_s = .305$, $t = 1.89$, $df = 34$, $.05<p<.10$). Because dependencies in the target letter sequence had proved (marginally) significant, however, we carried the analysis for letters a stage further even though there was no overall evidence of sequential dependency for the response letters. Specifically, we compared the signs of the deviations from expectation in target and response series to see whether any indication could be found that the subject's response pattern might have followed the marginally significant tendency for first order dependency between the target letters. In 25 cases the deviations are in the same direction, but in 38 cases in an opposite direction (one case was

dropped because one of the deviations was zero). This difference is not significant and even in an opposite direction from what would be expected if the subject had made use of the possible knowledge about the slight dependencies among successive target letters. Hence we can conclude that this marginally significant effect in the target series had no influence on the subject's response behavior, and that it can be neglected when analyzing the ESP data.

Since there was significant evidence of change (from first half to second half of the experiment) in van Dam's frequency of responding with the various letters (Table 9), we also checked to see whether these changes tended to be congruent with the significant changes that occurred across the two halves of the experiment in the target letter frequencies (Table 3). As can be seen from Table 11, the direction of the changes are the same for 7 of the 8 letters. However, this is nowhere near significant either by a sign test or by the more powerful Wilcoxon matched-pair signed-ranks test (Siegel, 1956). Thus there is no indication that van Dam was adapting his response pattern to match the shifting frequencies of the target letters. Moreover 3 of the 4 largest response shifts occur in directions congruent with his own overall biases, and contrary to those of the target series.

TABLE 11

Shifts in frequency for letters
between first and second half of experiment

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| targets | 3 | 26 | 2 | 1 | 5 | -21 | -6 | -9 |
| responses | -4 | 1 | 8 | 16 | 18 | -17 | -15 | -6 |

Finally, still assuming that van Dam was informed after each trial about the target for that trial, we investigated the relationship between target square and the response on the next trial to see whether we could find an indication that his responses were influenced by this possible knowledge of the target for the previous trial. No such relationship could be detected when applying a distance analysis between target and response on the next trial ($\chi^2 = 28.8$ df = 30).

To summarize, there is little evidence of overall congruence either in composition or structure of the sequences of targets and responses, and no evidence that van Dam's changing responses reflect increasing knowledge of properties of the target sequence. Although there are massive distortions in the overall frequencies of occurrence of the targets, there is little indication that van Dam's responses systematically reflect even these very gross distortions, and the inequalities of target and response frequencies are in any event easily handled in the analysis of direct hits through use of Stevens' method (Burdick and Kelly, 1977). In the statistically more critical area of sequential patterns, there is only slight evidence of dependency in each series separately, and less evidence of congruent biases between them. We therefore conclude that a valid statistical analysis of the experiment is entirely possible. In the sequel, this conclusion will receive further confirmation from several properties of van Dam's scoring performance. These include in particular the extremely high scoring rate; the fact that this scoring is widely distributed over the possible targets; and the fact that the scoring tends generally to be worse in places where the evidence of possibly relevant biases is, if anything, stronger--for example in the second half of the experiment, and in the numbers.

## Analysis of Direct Hits

Using the MDR response data (more distant responses, see page (263), van Dam scored 118 exact hits in 587 trials, approximately ten times the number expected by chance. This scoring is also widely distributed over targets; every number and letter target is hit substantially in excess of chance expectation, the letters in particular being hit at rates uniformly above 2.5 times expectation. Confusion matrices are presented in Tables 12 and 13. Evaluation of direct hits by Stevens' procedure leads to a z-score of 29.7 (MCE = 12.8, variance = 12.4). For the LDR data (less distance responses, see page (263) the same analysis leads to z = 32.1 ($N_{hit}$ = 127, MCE = 12.9, variance = 12.5).
We note in passing that the exact (Stevens) values for mean and variance differ only slightly from the theoretical binomial values (MCE = 12.23, variance = 11.97), or for that matter from the Poisson (MCE = variance = 12.23). Thus the choice of analysis model has very little impact on the enormous overall significance of the direct-hit results. In particular, the close correspondence between the binomial and Stevens results shows the minimal effect and uncorrelated character of the substantial distortions of the raw frequency distributions for targets and responses. The distance

TABLE 12

Confusion matrix for hits on letters

| | | Target A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|---|
| Response | A | 31 | 3 | 2 | 2 | 6 | 5 | 5 | 2 |
| | B | 10 | 32 | 5 | 0 | 4 | 6 | 5 | 2 |
| | C | 16 | 12 | 26 | 4 | 5 | 7 | 8 | 4 |
| | D | 11 | 15 | 13 | 22 | 11 | 7 | 11 | 2 |
| | E | 14 | 17 | 8 | 5 | 27 | 16 | 10 | 3 |
| | F | 8 | 8 | 10 | 4 | 14 | 26 | 15 | 0 |
| | G | 11 | 10 | 6 | 6 | 2 | 2 | 37 | 2 |
| | H | 2 | 4 | 3 | 0 | 4 | 2 | 7 | 8 |

TABLE 13

Confusion matrix for hits on numbers

| | | Target 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Response | 1 | 22 | 7 | 9 | 6 | 12 | 9 |
| | 2 | 8 | 33 | 13 | 5 | 18 | 6 |
| | 3 | 19 | 24 | 29 | 19 | 25 | 15 |
| | 4 | 20 | 26 | 31 | 39 | 27 | 16 |
| | 5 | 9 | 11 | 24 | 9 | 40 | 12 |
| | 6 | 7 | 12 | 4 | 4 | 4 | 11 |

analysis between targets and responses (see also page 260) shows not surprisingly that the observed distribution differs very strongly from the expected distribution, obtained by enumeration of all possible distances between all targets and all responses ( $\chi^2$ = 909 df = 30 p<<01). As can be expected, the main difference is caused by the excess of zero distances or hits.

The 233 trials in the distant condition by themselves are also highly significant, yielding 45 hits for z = 18.3 (MCE = 4.9, variance = 4.8). Furthermore, comparison of the distant versus near condition shows that there is no significant difference in scoring rate between these conditions, taking the individual squares as targets (Table 14).

TABLE 14

Hits and misses in distant and near conditions

| | near | distant | |
|---|---|---|---|
| miss | 281 | 188 | $\chi^2$ = .08 df = 1 |
| hit | 73 | 45 | |

Analogous results appear when analyzing the scoring in near vs. distant conditions for letters and numbers separately (letters: $\chi^2$ = 0.07, df = 1; numbers: $\chi^2$ = 0.2, df = 1). The trials in near and distant conditions also proved to be equally distributed over the three experimenters ($\chi^2$ = 0.4, df = 2); that is, each experimenter acted as agent in both conditions in relatively the same proportion of trials.

Comparing the scoring rates over the three experimenters, Heymans appears to be the most successful agent, and Brugmans the least successful. However, these differences among experimenters are not statistically significant (see Table 15; the unidentified fourth agent, who participated in 8 trials and obtained 1 hit, was omitted from this analysis).

TABLE 15

Scoring across experimenters

| | B | H | W | |
|---|---|---|---|---|
| miss | 189 | 132 | 141 | $\chi^2 = 3.4$ df = 2 |
| hit | 39 | 43 | 35 | |

Similar results are obtained in analyzing scoring over experimenters separately for letters and numbers, although in the scoring on letters there is a marginally significant difference ($p = .05$) mainly due to a slightly higher scoring rate when Weinberg acted as agent.

Not only is the rate of direct hits very high, but it is much higher than would be expected from chance association of hits on the letter and number attributes (see Table 16). This strongly confirms the conclusions tentatively reached by Foster (1952), based on data from the first seven sessions alone.

TABLE 16

Joint scoring on letters and numbers

| | miss number | hit number | |
|---|---|---|---|
| miss letter | 322 | 56 | $\chi^2 = 109.9$, df = 1, $p << .001$ |
| hit letter | 91 | 118 | |

From Table 16 we can also derive estimates of van Dam's scoring on the attributes separately, eliminating the direct hits. A simple and conservative way to do this is to analyze number hits only for trials in which letters are missed, and letter hits for trials in which numbers are missed. For letters, this reveals strong positive scoring with N = 413, MCE = 51.6, variance = 45.1, $Z \sim 5.7$. For numbers, however, there is a slight tendency toward missing (N = 378, MCE = 63.0, variance = 52.5, $Z < -1$).

As can be observed from Table 8, the subject showed much stronger zero order response effects on the numbers than on the letters. We also found stronger sequential dependencies in response numbers than in the response letter sequence. This suggests that the guessing aspect may have played a more important role in the up-down (number

attribute) direction than in the left-right direction. In other words, the left-right movements may have been dominant for the subject as compared to the movements in the up-down direction. Assuming that the excess of hits in those trials where no complete hit was achieved was due to the transfer of motoric impressions, this could explain the excess of hits on the letters. That is, it may have been easier for the experimenters to 'steer' the hand in the left-right direction, while in the up-down direction the subject was forced to rely more on guessing.

A further analysis of the scoring on the individual letters and numbers was carried out in which the expected scoring is based on the observed frequencies of targets and responses. That is, given the amount of hits on letters (209) we calculate, based on the relative frequencies of letters in target and response series, how this number of hits can be expected to be distributed over the individual letters. Comparing these expected values with the observed distribution of hits over the individual letters, we can test whether or not the scoring was significantly higher on some individual letters.

TABLE 17

Scoring on letters and numbers, adjusted for target and response frequencies

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| letter 0 | 31 | 32 | 26 | 22 | 27 | 26 | 37 | 8 |
| E | 27.4 | 31.5 | 28.8 | 18.8 | 35.4 | 28.0 | 35.9 | 3.3 |

$\chi^2 = 10.2$ df = 7 ns

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| number 0 | 22 | 33 | 29 | 39 | 39 | 11 |
| E | 16.5 | 27.5 | 43.1 | 38.9 | 38.3 | 8.7 |

$\chi^2 = 8.2$ df = 5 ns

Although the overall results of this analysis are entirely insignificant, there is a slight suggestion for both attributes that higher scoring rates occur on symbols which have lower expected values (due to their

low frequencies of occurrence as targets and/or responses). This trend is somewhat more pronounced for the numbers. Together, these results appear consistent with the suggestion that the subject did not treat the experiment as a guessing task, but that the guessing aspect may have played a more prominent role in the numbers or up-down direction. Also the scoring rate proved not be different for the left or right half of the board or for the upper or lower half.

## Analysis of Misses for Distance Effects

An important question in an experiment of this type is whether missing responses are still related in systematic though erroneous ways to the possible targets. Foster (1952) reports an analysis (based on data of Brugmans (1922), from the first seven sessions) directed at possible distance effects on letter and number attributes for responses which were not direct hits. The analysis reported seemed to suggest that missing responses might tend to occur in the physical neighborhood of the target. However, this analysis makes the untenable assumption that the targets were selected with equal probability.

We have approached the analysis of distance effects in two different ways. First, we performed complete consistent-missing analyses on the confusion-matrices for numbers and letters separately (see Tables 12 and 13), using the methods outlined in Burdick and Kelly (1977). A preliminary test for displacement effects showed that no such effects occurred for either attribute for +1 or -1 displacements. (Incidentally, the absence of -1 effects--choosing the target from the previous trial as the current response--is unusual, and suggests again that van Dam is not choosing responses in ways characteristic of subjects involved in guessing tasks with trial-by-trial feedback).

For numbers, the consistent-missing analysis is completely non-significant ($\chi^2_{CM}$ = 19.7, df = 19). For letters it is also non-significant ($\chi^2_{CM}$ = 47.7, df= 41), but there is a slight suggestion of a possible distance effect in that 10 of the 14 cells which are nearest-neighbors of the cells on the main diagonal have observed values exceeding their expectations, some substantially so.

Unfortunately, because of the small number of trials in relation to the size (48x48) of the relevant confusion matrix, the direct consistent-missing analysis cannot be extended to the individual target squares. It remains possible that statistically non-significant individual tendencies toward distance effects involving the separate attributes might also tend to occur jointly in the whole-target data.

To examine this possibility we carried out a distance analysis of the type described earlier, looking this time at the observed vs. expected distribution of distances between targets and responses, for misses only. The results are again entirely non-significant ($\chi^2 = 32.2$, df = 30). Examination of the detailed distributions (Figure 2) shows that the only visually interesting discrepancy occurs at the shortest distance, probably due to the non-significant excess hitting on nearest-neighbors for letters, mentioned above.

Thus the two analyses agree in showing that the scoring has been basically an all-or-nothing affair, with misses distributed across the board according to what would be expected from chance association of targets and responses.

## Analysis of Scoring Trends

We have analyzed for two classes of non-random temporal patterns in the distribution of scoring, namely decline effects and clustering of hits.

Figure 3 presents the distribution of the scoring rate over the sessions. It clearly shows a strong decline in scoring rate across the duration of the experiment. The experimenters felt that this decline was strong enough to justify the statement that van Dam had lost his telepathic ability. As has been stated in the description of the experiment, they evidently applied rather subjective criteria instead of statistical ones in forming this opinion. As can be seen from Table 18 the decline effect from first to second half of the sessions is strongly significant, but even in the last part of the experiment the scoring is still significant. We use a chisquare analysis in preference to the $CR_d$ because we are interested specifically in the question whether the observed scoring rates are different in the two halves of the experiment.

TABLE 18

Scoring in first and second half of the series

| | first 12 sessions | second 12 sessions | |
|---|---|---|---|
| miss | 278 | 191 | $\chi^2 = 16.0$ df = 1 $p<.001$ |
| hit | 94 | 24 | |

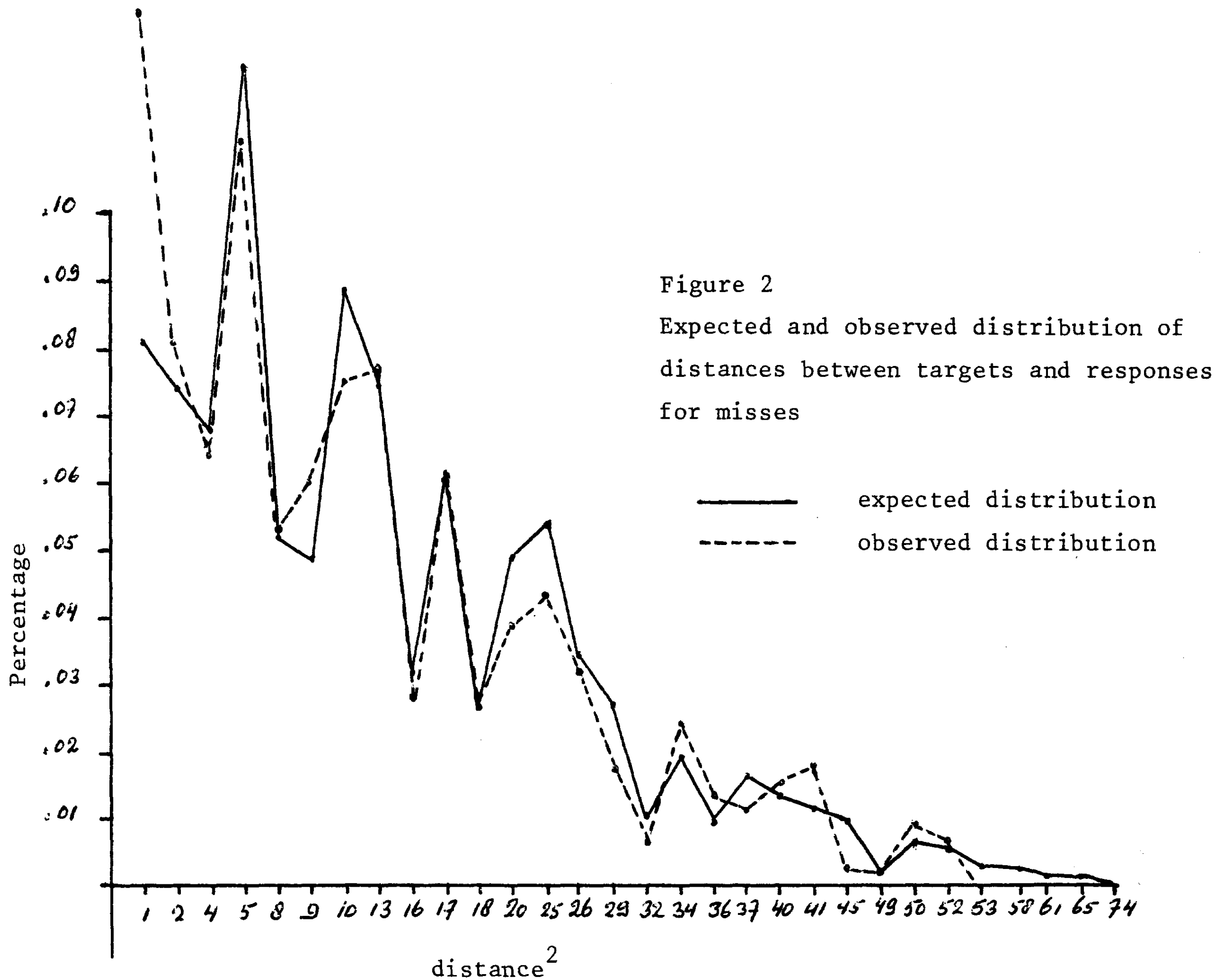



Figure 2
Expected and observed distribution of distances between targets and responses for misses

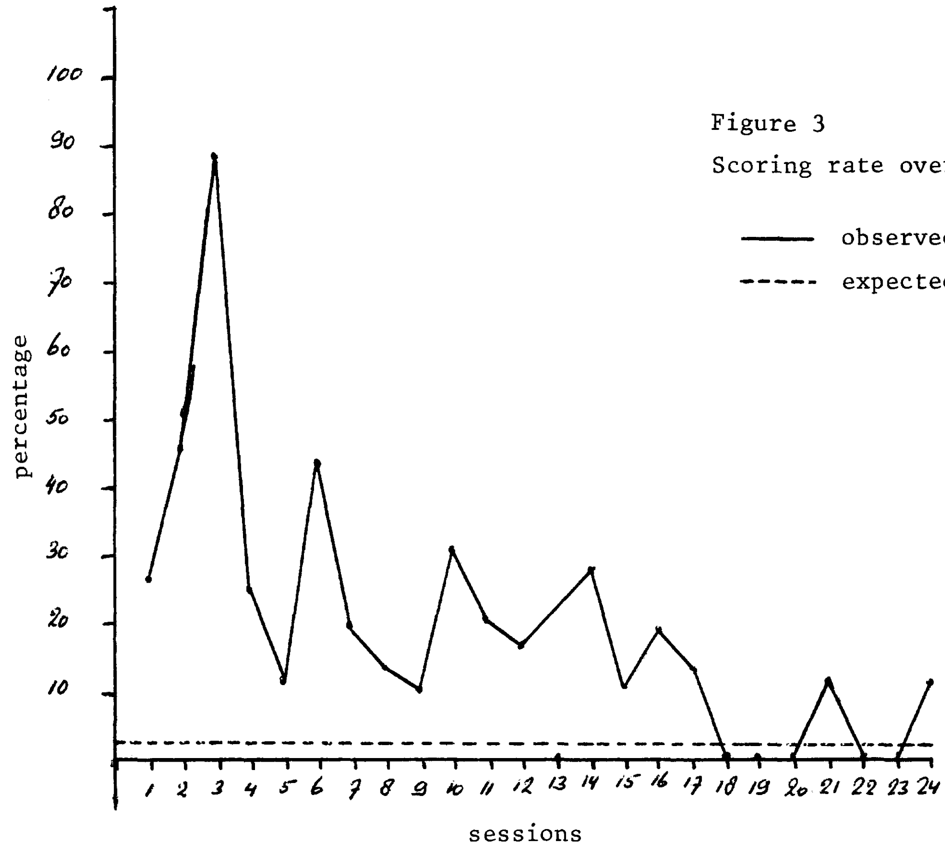



Figure 3

Scoring rate over the sessions

—— observed

----- expected

The last 5 sessions, held in May and June 1922, still yielded 7 hits in 83 trials ($p \sim .002$, Poisson). Nevertheless, if our assumption is correct that trial-by-trial feedback was provided, the strong overall decline seems to stand as a counter example to Tart's (1976) model of psi-stabilization. (The experimenters mention that even during the first seven sessions van Dam experienced increasing difficulties in getting into a suitable frame of mind for the experiments, but they attribute this difficulty to his increasing preoccupation with preparation for examinations toward the end of that period. No explanation is offered for the continuing decline in subsequent sessions, however).

Although the procedure is somewhat artificial because of the widely varying numbers of trials per session, we have also analyzed for within-session declines by dividing each session in half and pooling trials over sessions. Again using the chisquare analysis there is no evidence of differential scoring between session halves. First halves included 51 hits, second halves 66, $\chi^2 = 2.09$ with 1 df.

The experimenters recorded their opinion that van Dam scored most strongly when he had reached what they called a "passive" state. (Heymans, 1924). This opinion is based primarily upon physiological observations, but it is also consistent with the results of analyses for clustering of hits, independent of their overall rate of occurrence. Taking the entire series as a unit, the Wald-Wolfowitz z-score for strings (Burdick and Kelly, 1977) is -5.16, $p < .001$. (Although the procedure is again somewhat artificial, this time because it ignores run boundaries, this circumstance appears unlikely to affect the result or its interpretation; only one hit-string crosses a run boundary, a string of length 2 connecting sessions 4 and 5). Similar results are obtained for letters ($z = -2.85$, $p < .003$) and for numbers ($z = -5.68$, $p < .001$), although these are of course not independent of the main result for squares. Two sessions contributed especially strongly to the stringing effect; session 2, in which 11 hits were scored in 24 trials including a string of 8 hits; and session 3 in which there were 16 hits in 18 trials, including a string of 12 (furthermore, the two misses were on adjacent squares). Even in the less successful later sessions, however, the hits still tend to cluster even though the groups are smaller and more scattered. Thus, whatever conditions were conducive to success in the task tended, once established, to persist for periods of time sufficient to span a number of trials. Furthermore, this effect is certainly underestimated due to the form of the presently

available statistical tests, which in effect dichotomize the data by considering only one degree of success at a time. (To illustrate, in the example given above for session 3, van Dam received no partial credit for his two near-misses. The exact-hit stringing analysis thus counts five strings in that session, where a more sensitive procedure might choose a number much closer to one).

## Analyses of Response-Time Data

The overall distribution of response times follows a characteristically chisquare-like pattern, with considerable right-hand skew caused by small numbers of very slow responses. Fur purposes of analysis we have therefore either applied a normalizing (log) transform (for ANOVA), or blocked the overall distribution into a small number of bins (for table analysis).

Both forms of analysis indicate a significant association between response time and distance from the target, which results from an excess of exact hits at short response times (see Figure 4). The simple 1-way ANOVA for misses vs. hits on squares has ${}_1F_{368}$ = 13.5, $p \leq .001$, and the same pattern appears for numbers and letters separately, though with lesser strength. Furthermore, response time is influenced by other factors as well: for example, there is a massive difference between near and distant conditions (${}_1F_{568}$= 36.8, $p \ll .001$), the trials in the near condition being completed faster; and there is a marginally significant difference in the distribution of response times within the two halves of the sessions ($\chi^2$ = 20.1, df = 10, $p < .05$), responses tend to be slower in the second half. In both cases no difference exists as regards scoring rate, and so the explanation appears likely to involve normal psychological factors. The slower responses on distant trials, for example, may result from van Dam's feeling less confident and more hesitant under these conditions. Likewise the slowing of responses over the course of a session may reflect some sort of decline in motivation.

In light of the above, we reanalyzed the relationship between hitting and response time as a two-way ANOVA, extracting the variance associated with the near/distant factor. There is no interaction, and the main effects for both factors persist at their original levels of significance or better. Moreover neither effect appears artifactually dependent on a hidden relationship between response time and series position (which is correlated with both miss/hit and near/distant).

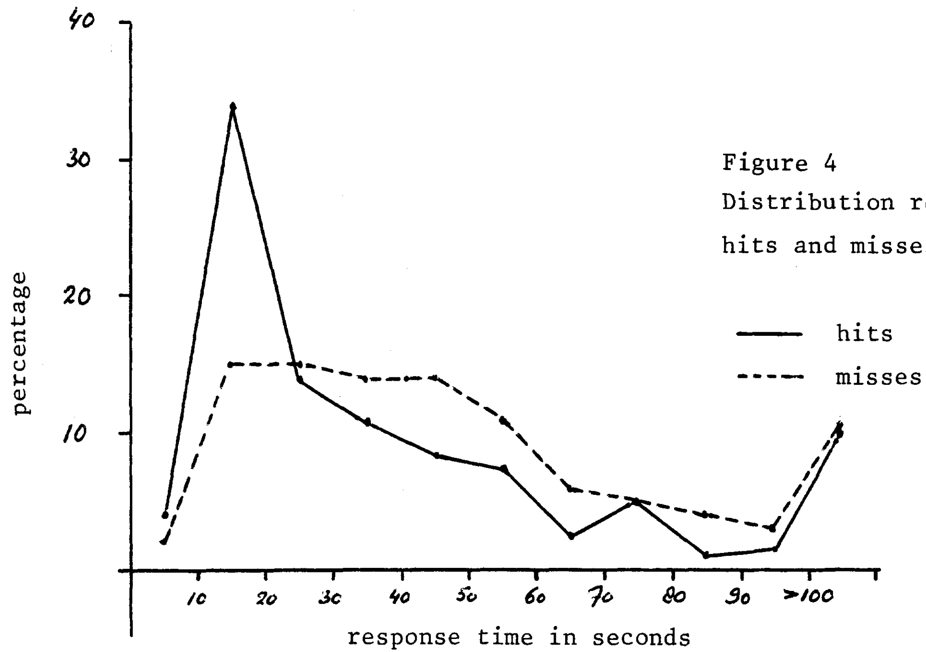



Figure 4
Distribution response times for
hits and misses

Thus we have clear evidence that correct responses tend on the average to be faster. (At the same time, however, we should emphasize that a number of very slow responses were also hits, including the slowest response of the entire experiment--9 minutes, 9 seconds. Indeed, the score rate is above chance at every level of response time in Figure 4). There may be numerous factors influencing response time, and the interpretation of this result does not appear straightforward. Unfortunately, the available reports contain no helpful information on this question.

A further puzzle is raised by an apparent effect of experimenters on response time (${}_{2}F_{559} = 5.83$, $p < .003$), contributed mainly by slow responses occurring when Brugmans acted as agent. We unfortunately cannot tell from the reports whether van Dam typically knew in advance the identy of the agent, so that this result might or might not represent some sort of psi effect.

The Ambiguous Responses

In 131 trials an ambiguous response was recorded. As stated before, all analyses presented have been based on the MDR set of responses, using those responses in the ambiguous trials which were more distant from the target square.

As mentioned above, neither the target letter distributions nor the target number distributions differ significantly for ambiguous and non-ambiguous trials (letters: $\chi^2 = 8.8$, df = 7; numbers: $\chi^2 = 4.4$, df = 5).

The distribution of response letters and numbers in the LDR set of responses do differ for ambiguous and non-ambiguous trials (letters: $\chi^2 = 14.3$, df = 7, p = .05; numbers: $\chi^2 = 21.6$, df= 5, $p < .01$), but these differences disappear when the distributions for letters and numbers of the MDR set of responses are analyzed (letters: $\chi^2 = 10.8$, df = 7; numbers: $\chi^2 = 9.3$, df = 5). We assume that the deviant distributions in the LDR set of responses are caused by the way these responses were formed. We know that the targets were not equally distributed over the board, and in particular that the numbers tended to be concentrated in the centre. By taking systematically those responses which are closest to the target we probably introduce a slightly deviant frequency distribution for those responses, in effect exaggerating van Dam's already pronounced tendency to choose responses in the interior of the checkerboard. This is directly supported by examining the LDR and MDR response frequency distributions. For both letters and numbers the LDR set shows higher frequencies in the interior, and in particular the significant non-uniformity of the number response

distribution in the LDR set of responses is mainly caused by strong negative deviations for the extreme numbers, that is the numbers 1 and 6. Thus the MDR dataset appears in general to represent a more "natural" way of resolving the response ambiguities for purposes of statistical analysis, although as we have indicated none of the main results is affected by this choice.

We have not been able to determine unequivocally the sources of response ambiguity, which could lie either in the subject, the observers, or both. Before discussing these possibilities, let us outline the main facts which a correct account must explain.

First, the scoring rates for trials with ambiguous responses are markedly lower than for trials with clear responses. Even using the LDR data (since the MDR set has by definition no exact hits on ambiguous trials), the effect is strongly present in the scoring on exact squares (Table 19).

TABLE 19

Scoring on squares in ambiguous and non-ambiguous trials

LDR set of responses

| | non-ambiguous | ambiguous | |
|---|---|---|---|
| miss | 338 | 122 | $\chi^2 = 20.6$, df = 1, $p < .001$ |
| hit | 118 | 9 | |

In this table the hit rate for non-ambiguous responses is almost 4 times that for ambiguous responses, even though the latter is itself better than 3 times MCE, $p < .003$( Poisson). For the numbers the same effect is marginally present ($\chi^2 = 4.4$, df = 1, $p < .05$), and for the letters it is again very strong ( $\chi^2 = 18.9$, df = 1, $p < .001$).

Second, the response times are markedly longer for the ambiguous trials, by both a chisquare analysis ( $\chi^2 = 25.3$, df = 10, $p < .005$) and the ANOVA using log response time ( $_1F_{568} = 9.76$, $p < .002$). There is also a significant association between experimenter and clear vs. ambiguous response due mainly to a large excess of ambiguous responses occurring with Brugmans as agent (Table 20), and as mentioned earlier the response times for Brugmans are also longer; however, the appropriate 2-way ANOVA's show that these effects are independently significant and additive.

TABLE 20

Response ambiguity vs. experimenters (MDR data)

| | B | H | W | |
|---|---|---|---|---|
| clear | 163 | 139 | 146 | $\chi^2 = 8.06$, df = 2, $p < .02$ |
| ambiguous | 65 | 36 | 30 | |

Finally, we looked at the form of the response ambiguities, to determine whether they occurred predominantly in one or the other direction. The difference as regards direction is non-significant, considering that there are seven boundaries for letters and five for numbers. (Table 21).

TABLE 21

Form of response ambiguity

| letter differs, number equal | number differs, letter equal | letter and number differ |
|---|---|---|
| 54 | 42 | 35 |

What seems to us the most plausible interpretation of those observations requires as an essential premise that van Dam could distinguish (presumably by tactile means) the boundaries between squares. The reports do not explicitly state that this was the case, but the photographs suggest that the lines were etched or dug into the surface of the board. We then suppose that van Dam himself was the main source of ambiguity, deliberately tapping between two squares when he felt uncertain (and hence was more likely both to respond slowly and to end by missing the target). The higher rate of ambiguous responses with Brugmans acting as agent would be easy to explain if van Dam knew the agent for each trial in advance and simply felt less confident with Brugmans. Even if he did not consciously know the agent, the result might still be a psi effect indicating that he knew unconsciously, or that Brugmans was for unknown reasons a less effective agent. The view that van Dam was the source also appears relatively consistent with our knowledge of the physical layout of the situation and with the data on the form of ambiguous responses: From the photographs of the checkerboard taken down through the ceiling it appears very unlikely that the

observers would have had any difficulty in correctly reading an unambiguous response by van Dam.

Plausible though it is, this is only one of possible "stories" about the data, and the available material is not sufficient to allow us to choose decisively among them. In particular it remains quite possible that ambiguity was contributed from the observer's side, either solely or in addition to ambiguity introduced knowingly or unknowingly by van Dam. For example, the excess ambiguity when Brugmans was agent might indicate that Heymans and Weinberg (who in this case functioned as the observers) tended to be more indecisive recorders and more inclined to rate a given response as ambiguous (of course we would then have to explain why they chose systematically thus to treat responses which were slower and more likely to be misses, but perhaps these factors were also related to some aspect of the trajectory of van Dam's hand that made it more difficult to read his final response).

Another possibility is that the observers knew what the target was, and tended to stay on the safe side by rating ambiguous trials which were potential hits as non-ambiguous misses. Alternatively, they might have had a tendency to rate some ambiguous trials as complete hits when they had a choice between the target square and an adjacent square. Both explanations suppose, however, that the observers knew the target before the response was recorded. They might conceivably have learned this from the behavior of the agent, but in that case we should expect a strong difference between the near and distant condition as regards the amount of ambiguous trials. In the distant condition the agent stayed in the upper room, could not be seen or heard by the subject and hence he was rather free to show his expressions. It is explicitly stated, however, that to avoid sensory cues the agent tried to control himself as much as possible in the near condition. However, the difference as regards number of ambiguous responses between near and distant condition is marginally significant in favor of the latter, the distant condition having relatively more ambiguous trials (see Table 22). If the observers had tried to guess the

TABLE 22

Ambiguous responses in near and distant conditions

| | near | distant | |
|---|---|---|---|
| non-ambiguous | 285 | 170 | $\chi^2 = 4.45 \quad p<.05$ |
| ambiguous | 68 | 63 | |

target from the agent's behavior in order to avoid ambiguity when the response was close to the target area, one should expect an opposite effect, because it must have been much more difficult for them to guess the target in the near condition. The results of Table 22 make much more sense in terms of our original hypothesis, that van Dam voluntarily generated ambiguous responses when he felt particularly uncertain.

If we assume the worst, that is, that the observers knew the target and tended to rate ambiguous responses bordering the target square as complete hits, we can calculate an estimate of the resulting excess of complete hits based on the ratio of ambiguous versus non-ambiguous responses for the misses. This ratio is 122:338 (see Table 19). There are 118 non-ambiguous responses which are complete hits. We now assume that part of these hits should have been recorded as ambiguous responses. Hence the number of 118 is reduced by 122/338 x 118 - 9 = 34, we subtract 9 because 9 ambiguous hits have been recorded. So based on these assumptions we should have expected approximately 84 direct hits and 43 ambiguous hits. Even in that case the result of the experiment would be highly significant when considering complete hits only, while for the distance distributions between target and responses the difference would be minimal.

DISCUSSION

The experiment and its statistical results have been presented in considerable detail. To complete the agenda set forth in our introduction we will now draw upon the results outlined above in reassessing the question of possible fatal methodological difficulties in the experiment. The statistical difficulty, that the properties of the target order might be inadequate to support valid statistical analysis of the data, has been decisively overthrown. There can be no reasonable doubt that results of extreme statistical significance were obtained. The procedural difficulty, however, that these results might be explained by sensory leakage, is more serious and requires further discussion.

The hypothesis advanced by Soal and Bateman (1954) and eagerly accepted by Hansel (1966), is that the observers may have become excited as van Dam's hand approached the target square, and hence quite unintentionally (perhaps through some sort of bodily movement shifting the distribution of weight on the floorboards) conveyed an auditory "stop" signal.

Whately Carington (1946) had considered and rejected this same argument. He had paid a visit to Groningen in 1937 and carefully

examined the floor between the two rooms, which he found to be "... of such solid construction that no ordinary movement, change of breathing, or the like, could possibly have acted as a "stop" signal unless we assume a quite extraordinary degree of hyper-sensitivity on the part of the subject..." Soal was willing to make that assumption, and rationalized it by pointing to the case of the stage-telepathist Fred Marion as an example of the required sensitivity. Although his discussion acknowledges that the experimenters were good psychologists and likely to be on guard against auditory cues, Soal clearly intimates that they would not have been prepared by experience to cope with the capacities of a Fred Marion.

Evidently Soal was not familiar with the background of the experiment, which in this instance is essential. Not only were the experimenters prepared in a general way to deal with problems of sensory leakage, but also they were prepared specifically for the problems posed by a Fred Marion, or in their case a Rubini. They were entirely aware of the possibility that sensory cues not noticeable to "normal" people might produce hits, and state explicitly their view that all of the Rubini-type demonstrations were defective for that reason. Against this background, their procedures were specifically framed so as to be "Rubini-proof". The elimination of visual cues can be unambiguously judged as completely satisfactory, based upon the reports and the photographs. The difficulty in judging the more critical auditory case arises because the reports do not and cannot convey in an analogously simple way the adequacy of the physical barrier. But certainly the experimenters' testimony that they took all possible prior precautions to eliminate auditory cues, in full awareness of the Rubini problem, must be taken seriously.

It also seems likely that in practice the experimenters would quickly have detected any general contingency of the hypothesized sort between van Dam's success and their behavior upstairs. Scrapes or thumps on the floor audible downstairs to van Dam would almost certainly have been audible to the experimenters as well.

Furthermore, such cues even had they existed would probably not have been adequately specific to identify the target. Recall that the agent's general task was constantly to steer van Dam's hand toward the target. Rather than passively awaiting the final outcome, the agent was involved throughout the course of a trial and hence if generating any auditory cues at all, probably generating many cues of similar form. Also the observers might have generated cues unrelated to the target, particularly at the beginnings of trials.

The auditory-cues hypothesis also leads to several other expectations that seem inconsistent with the data, even if we allow that target-specific cues could have occurred as van Dam groped over the target square. First, one would presumably expect to find a statistical excess of responses in near-neighborhoods of the target squares, but this effect did not occur. Second, the scoring might reasonably be expected to differ across experimenters because of subtle differences in the subtle cues they produced in playing the role of agent, but it does not. Similarly, scoring might be expected to differ between near and distant conditions because of differences in quality and quantity of available cues, but it does not. Third, one might expect that longer trials would be more likely to result in hits, since on the average van Dam's hand would cross the target area more often and thus provide more opportunities for him to receive auditory cues; exactly the reverse occurred. This argument assumes in fact a constant generation of very weak auditory cues. In the case of occasionally clear auditory cues, an opposite effect could be expected. The subject could then respond quickly with a good chance of being correct, while in those trials where no cues were provided he would tend to take more time for his response and would be more likely to miss the target. But considering the construction of the floor between the two rooms, it is difficult to imagine that such strong auditory cues could have been perceived by the subject but would have passed unnoticed by agent and observers. Finally, one could reasonably expect van Dam to improve or at least maintain his ability, but instead it declined rather swiftly.

Therefore, although absolute certainty cannot be claimed, the weight of the historical and experimental evidence seems to us to favor strongly the experimenters' own belief that they had succeeded in eliminating sensory alternatives to the psi hypothesis. Carington (1938) stated that he had carefully discussed the possibility of 'stop signals' with Prof. Brugmans who however remained convinced that nothing of the kind could have occurred. Both the principal investigators, Heymans and Brugmans, became convinced of the reality of telepathy by this experiment, amongst others because they felt they had excluded all possible sensory cues. There was certainly room for improvement in the randomization and recording procedures, and the reports could easily have eliminated the need for several inconclusive discussions merely by giving more details about procedural matters such as the physical structure of the checkerboard, the sources of response ambiguity, and the information van Dam received before and after trials. Nevertheless the experiment remains in our view professionally done and

fundamentally sound, in many respects still very good and certainly exceptionally good for its time.

It also contains a rich variety of interesting and statistically strong effects, a number of which are reported here for the first time. The results that strike us as most relevant to contemporary needs in parapsychology, however, are those such as the stringing results and most particularly the GSR results-- suggestive of a psi-conducive psychophysiological state. Certainly one of the most conspicuous facts about psi phenomena is their tendency to congregate around certain individuals. Correspondingly, a high-priority task for experimenters is to press as far as possible analysis of the factors at work in such individuals, that may account for their exceptional abilities in this kind of research. The experiment of Brugmans, Heymans, and Weinberg is certainly a major landmark.

ACKNOWLEDGEMENTS

Parts of this study have been presented at the 1978 PA Convention by E.F. Kelly.

REFERENCES

Brugmans, H.J.F.W. "L'Etat passif" d'une télépathe, controlé par le phénomène psychogalvanique. L'Etat Actuel des Recherches Psychiques d'après les Travaux du II me Congrès International, Paris, 1924, 95-125.

Brugmans, H.J.F.W. De "Passieve Toestand" van een telepaath, door het psychogalvanisch phenomeen gecontroleerd. Meededeelingen der Studievereeniging voor Psychical Research, 7, 1923, 65-124.

Brugmans, H.J.F.W. Une communication sur des expériences télépathique au laboratoire de psychologie à Groningen faites par M. Heymans, Docteur Weinberg et Docteur H.J.F.W. Brugmans. Le Compte Rendu officiel du Premier Congrès International des Recherches Psychiques, Copenhague, 1922, 396-408.

Burdick, D.S., Kelly, E.F. Statistical methods in parapsychological research. In Wolman, B.(ed.) Handbook of Parapsychology, New York: Van Nostrand Reinhold, 1977, 81-130.

Carington, W. Some early experiments providing apparently positive evidence for extra-sensory perception. J.S.P.R., 30, 1938, 299-302.

Carington, W. Thought Transference, New York: Creative Age Press, 1946, 37-39.

Davis, J.W., Akers, C. Randomization and tests for randomness. Journal of Parapsychology, 4, 1974, 393-407.

Foster, E.B. Multiple-aspect targets in tests of ESP. Journal of Parapsychology, 16, 1952, 11-22.

Hansel, C.E.M. ESP: A Scientific Evaluation. New York: Charles Scribner's Sons, 1966, 35-36.

Heymans, G., Brugmans, H.J.F.W., Weinberg, A.A. Een experimenteel onderzoek betreffende telepathie. Meededeelingen der Studievereeniging voor Psychical Research, 1, 1921, 3-7.

van Loon, F.H. A method of investigation into thought-transference. J.S.P.R., 1921, 4-23; 34-49.

Murphy, G. Telepathy as an experimental problem. In Murchison, C. (ed.). The Case for and against Psychic Belief. Worcester, Mass: Clark University, 1927, 268-270.

Murphy, G. Challenge of Psychical Research. New York: Harper Brothers, 1961, 56-62.

Pope, D.H. The Brugmans experiments. Journal of Parapsychology, 16, 1952, 1-3.

Rhine, J.B. History of experimental studies. In Wolman, B.(ed.) Handbook of Parapsychology, New York: van Nostrand Reinhold, 1977, 25-47.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Soal, S.G., Bateman, F. Modern Experiments in Telepathy. London: Faber and Faber, 1954, 15-16.

Tart, C.F. Learning to Use Extrasensory Perception. Chicago: University of Chicago Press, 1976.